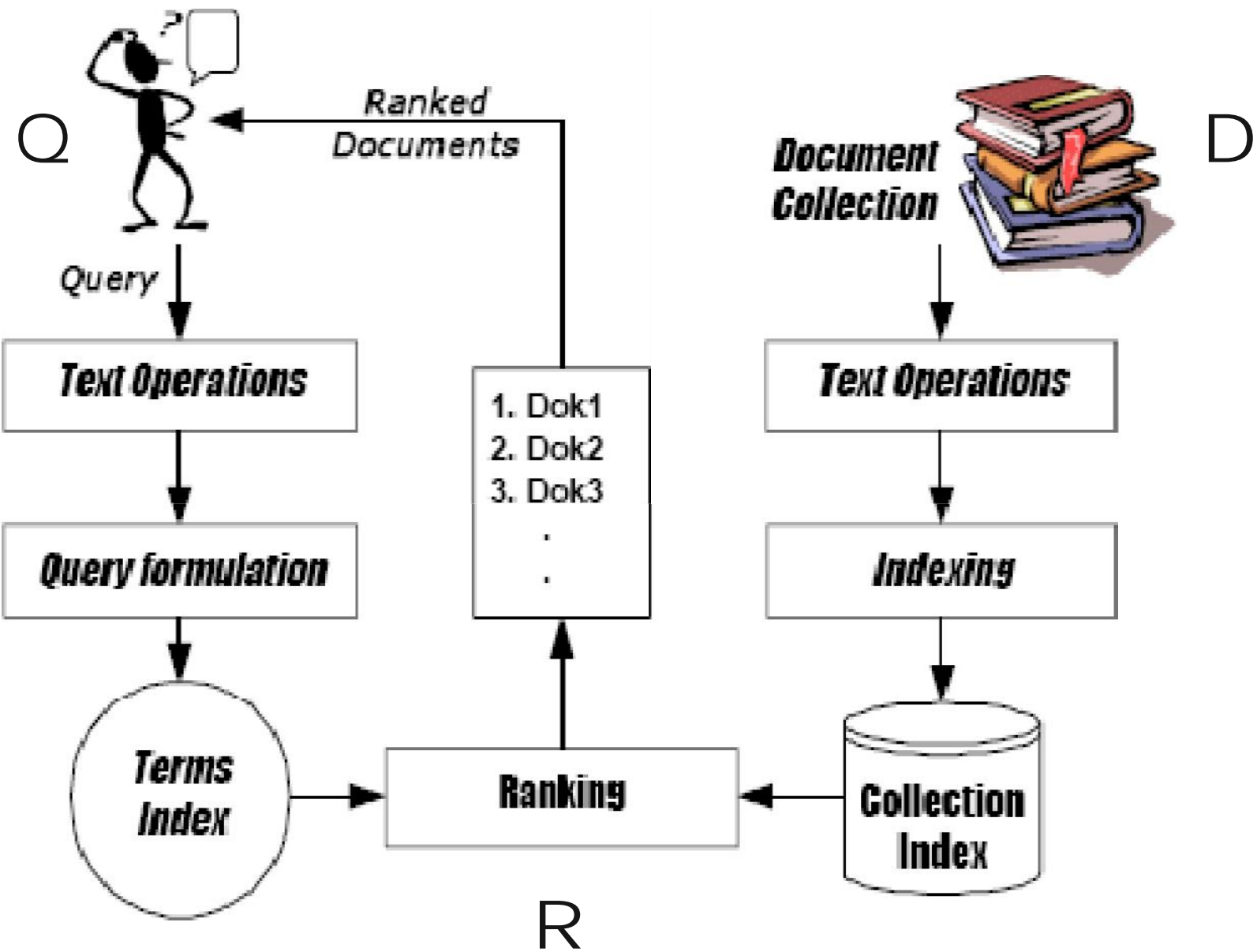




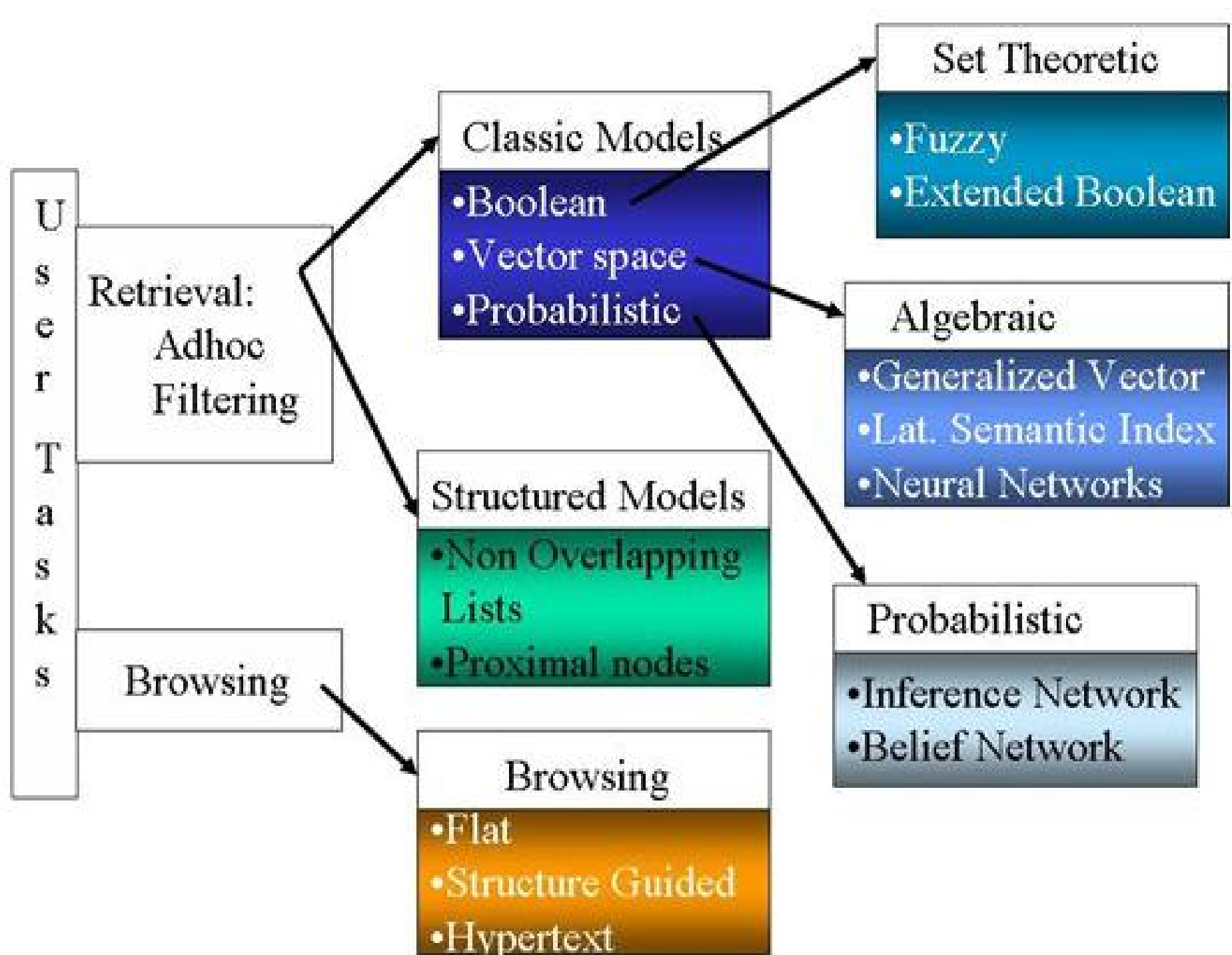
MODEL
TEMU BALIK
INFORMASI

SUGENG PRIYANTO



Formal Karakteristik Sistem TB

- ❖ Himpunan yang terdiri dari logical view bagi koleksi-koleksi dokumen $\rightarrow D$
 - ❖ Himpunan yang terdiri dari logical view bagi kebutuhan informasi user. Representasi tersebut disebut query $\rightarrow Q$
 - ❖ Framework bagi representasi dokumen, query, dan hubungannya $\rightarrow F$
 - ❖ $R(q_i, d_j)$ adalah fungsi rangking (peringkat) yang berasosiasi antara bilangan real dengan query $q_i \in Q$ dan representasi dokumen $d_j \in D$. Rangking tersebut mendefinisikan urutan di antara dokumen-dokumen yang diambil dengan menggunakan query q_i
-



3 MODEL KLASIK

LOGO

- ❖ Model Boolean
- ❖ Vector Space Model
- ❖ Probabilistic

<http://iperpin.wordpress.com/category/information-retrieval/page/3/>

Model Boolean

- ❖ Merupakan model yang paling sederhana
 - ❖ berdasarkan teori himpunan dan aljabar Boolean
 - ❖ Model ini menggunakan operator boolean yaitu AND, OR atau NOT
 - ❖ Metode ini merupakan metode yang paling sering digunakan pada mesin penelusur (search engine) karena kecepatannya
-

Teori Himpunan

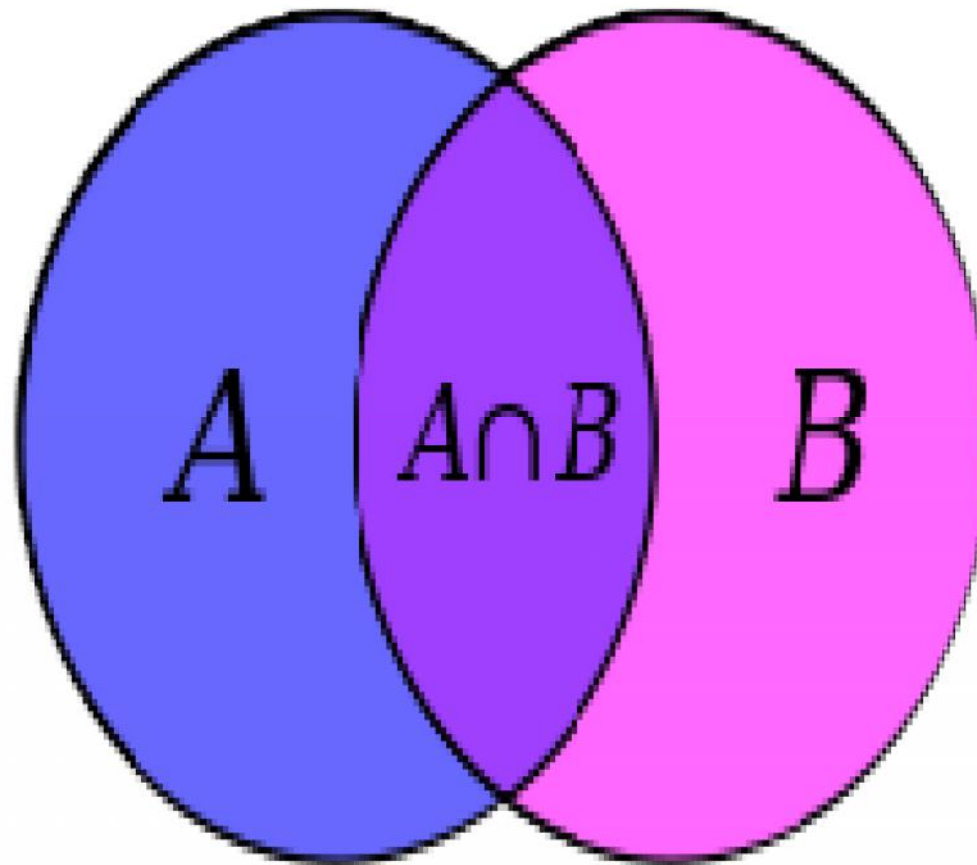
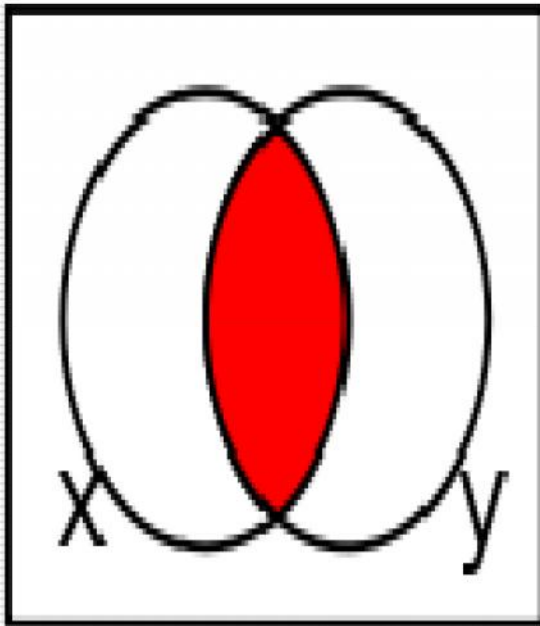
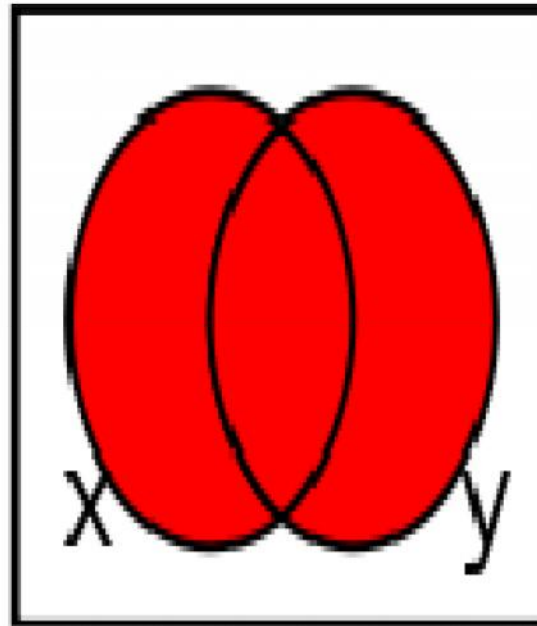


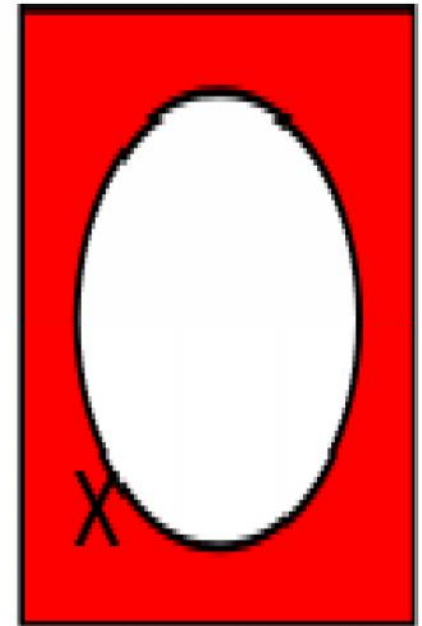
Diagram Vena Boolean



$$x \wedge y$$



$$x \vee y$$



$$\neg X$$

Tutorials

LOGO

<http://lib.colostate.edu/tutorials/boolean.html>

Keuntungan model Boolean

- ❖ merupakan model sederhana sehingga mudah diimplementasikan
 - ❖ Berguna bagi pengguna yang tidak memiliki pengetahuan banyak mengenai formulasi kueri
 - ❖ Efisien
-

Kerugian

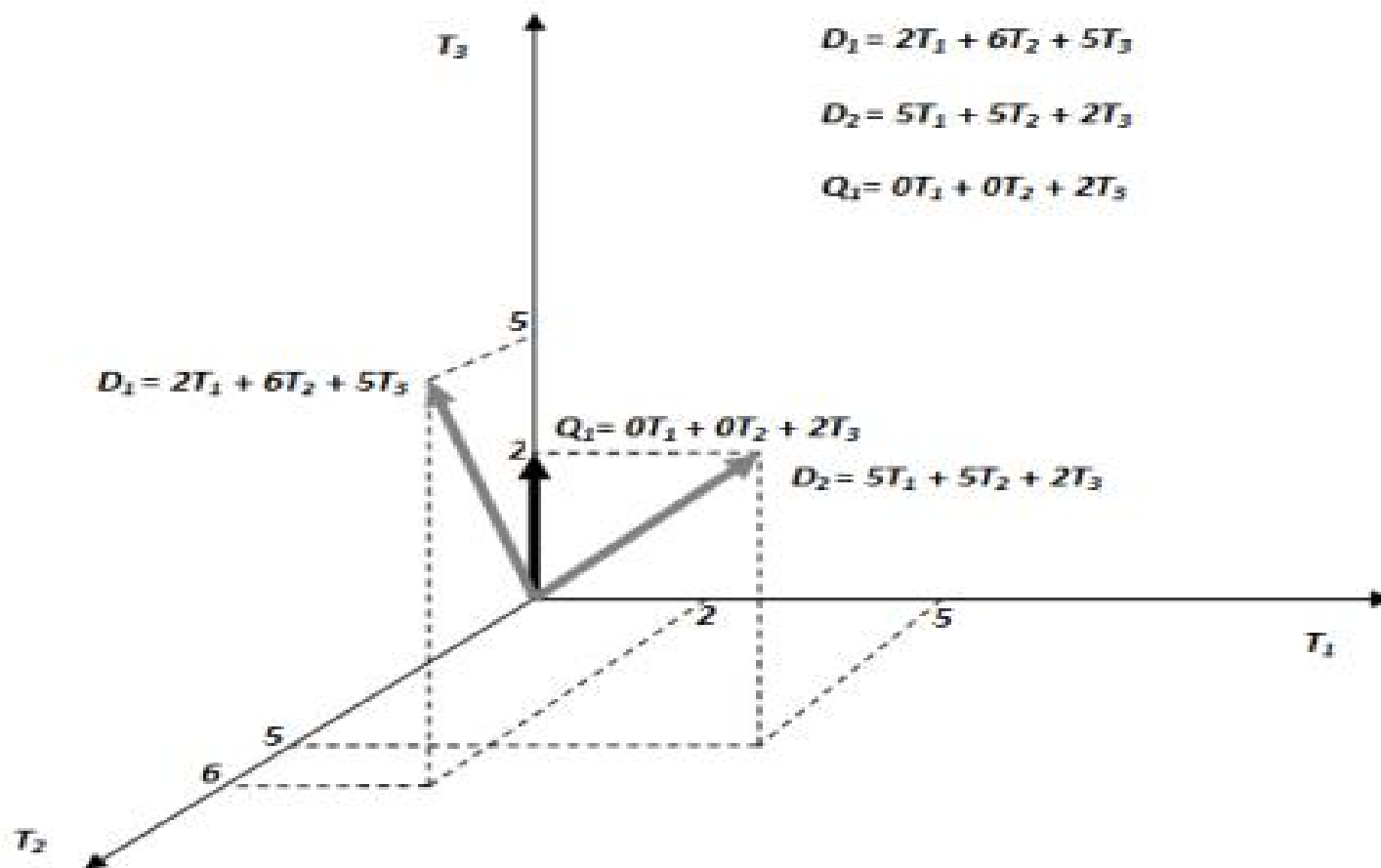
- ❖ Tidak ada peringkat dokumen (ranking)
 - ❖ Tidak mendukung partial matching
 - ❖ Secara keseluruhan tergantung kepada definisi keyterm dan pembobotan
 - ❖ Dokumen yang terambil hanya dokumen yang benar-benar sesuai dengan pernyataan boolean/kueri yang diberikan
-

Model Ruang Vektor

- ❖ Model vektor berdasarkan keyterm
 - ❖ Model vektor mendukung partial matching dan penentuan peringkat dokumen
 - ❖ Model vektor memerlukan :
 - Bobot keyterm untuk vektor dokumen
 - Bobot keyterm untuk kueri
 - Perhitungan jarak untuk vektor document-keyterm
-

Konsep

- ❖ Model ruang vektor dibuat berdasarkan pemikiran bahwa isi dari dokumen ditentukan oleh kata-kata yang digunakan dalam dokumen tersebut.
 - ❖ Model ini menentukan kemiripan (similarity) antara dokumen dengan query dengan cara merepresentasikan dokumen dan query masing-masing ke dalam bentuk vektor.
 - ❖ Tiap kata yang ditemukan pada dokumen dan query diberi bobot dan disimpan sebagai salah satu elemen vektor
-



Gambar 2.5 Contoh Model Ruang Vektor dengan dua dokumun D_1 dan D_2 , serta *query* Q_1 . [1]

Keuntungan

- ❖ Efisien
 - ❖ Mudah dalam representasi
 - ❖ Dapat diimplementasikan pada document-matching
-

Kerugian

LOGO

- ❖ Teoritical Frameworknya tidak jelas
 - ❖ Menghasilkan indeks yang berdekatan
 - ❖ Asumsi yang digunakan adalah independensi index term
-

MODEL PROBABILITAS

- ❖ model probabilistik mengasumsikan bahwa setiap dokumen dideskripsikan lewat "ada" atau "tidak ada"nya term indeks
 - ❖ Menggunakan pendugaan probabilistik untuk menentukan dokumen yang relevan dengan keyterm yang diberikan
 - ❖ Cukup kompleks
 - ❖ Kinerjanya lebih efisien dibandingkan dengan model ruang vektor
 - ❖ Probability : Menggunakan semantik
-

